Lectures 2 and 3

Covariance between two random variables. X and Y.

$$Cov(X,Y) = E[[X - E[X]][Y - E[Y]]] = E[XY] - E[X]E[Y]$$

If $X = F(x)$ and $Y = G(x)$, then

$$Cov(X,Y) = \sum_x F(x)G(x)p(x) - \sum_x F(x)p(x)\sum_x G(x)p(x)$$

In particular

$$\begin{aligned} Var(\sum_i X_i) &= \sum_i Var(X_i) + \sum_{i\neq j} Cov(X_i, X_j) \\ &= \sum_i Var(X_i) + 2\sum_{i>j} Cov(X_i, X_j) \end{aligned}$$

We have a population of size N. We want to know what percentage support a certain government policy. We select a sample of size n from N. Suitably randomly chosen. Every subset of size n has probability $\binom{N}{n}^{-1}$ of being chosen. The probability that a particular individual is chosen is

$$\frac{\binom{N-1}{n-1}}{\binom{N}{n}} = \frac{n}{N}$$

The probability that everyone in a specified subset of r individuals is selected is

$$\frac{\binom{N-r}{n-r}}{\binom{N}{n}} = \frac{n(n-1)\cdots(n-r+1)}{N(N-1)\cdots(N-r+1)}$$

Probability that a specified person is selected is $\frac{n}{N}$. Probability that two given individuals are selected is $\frac{n(n-1)}{N(N-1)}$.

Suppose that out of N, M support the policy. We want to know the value of $p = \frac{M}{N}$. We select a sample of size n and see how many support the policy. If that is m then we offer $X = \frac{m}{n}$ as an estimate of p. We need to compute $E[X]$ and $Var(X)$. Let $\{i\}$ be an enumeration of the individuals. $\epsilon(i) = 1$ if the i-th individual supports the policy and 0 otherwise. What we need is to estimate $\frac{1}{N}\sum_i \epsilon(i)$. Let $\eta(i) = 1$ if i-th individual is included in the sample and 0 otherwise. $n = \sum_i \eta(i)$ and $m = \sum_i \epsilon(i)\eta(i)$.

$$X = \frac{1}{n}\sum_i \epsilon(i)\eta(i)$$

$E[\eta(i)] = \frac{n}{N}$, $Var[\eta(i)] = \frac{n}{N}(1-\frac{n}{N})$,

$$Cov(\eta(i)\eta(j)) = \frac{n(n-1)}{N(N-1)} - \frac{n^2}{N^2} = -\frac{n}{N}\left[\frac{n}{N} - \frac{n-1}{N-1}\right] = -\frac{n}{N}\frac{N-n}{N(N-1)}$$

$$E[X] = \frac{1}{n}\sum_i \epsilon(i)E[\eta(i)] = \frac{n}{N}\frac{1}{n}\sum_i \epsilon(i) = \frac{M}{N} = p$$

$$\begin{aligned} Var(X) &= \frac{1}{n^2}\sum_i \epsilon(i)Var[\eta(i)] + \frac{1}{n^2}\sum_{i\neq j}\epsilon(i)\epsilon(j)Cov[\eta(i)\eta)j] \\ &= \frac{M}{n^2}\frac{n}{N}(1-\frac{n}{N}) - \frac{M(M-1)}{n^2}\frac{n}{N}\frac{N-n}{N(N-1)} \\ &= \frac{1}{n}(1-\frac{n}{N})\left[\frac{M}{N}\right]\left[1-\frac{M-1}{N-1}\right] \\ &\simeq \frac{p(1-p)}{n} \end{aligned}$$

Standard Deviation $\sigma(X) = \sqrt{Var(X)}$.

Tchebychev's inequality.

$X \geq 0$. $E[X] \geq \ell P[X \geq \ell]$. $P[X \geq \ell] \leq \frac{E[X]}{\ell}$.

$$\begin{aligned} P[|X - E[X]| \geq k\sigma(X)] &= P[|X - E[X]|^2 \geq k^2 E[X - E[X]|^2] \\ &\leq \frac{E[X - E[X]|^2}{k^2 E[X - E[X]|^2} \\ &= \frac{1}{k^2} \end{aligned}$$

$X \pm k\sigma(X)$ will cover $E[X]$ with at least $1-k^{-2}$ level of assurance. $k=10$, 99% sure! But in practice k can be much smaller.

Law of large numbers. If $X_1, \ldots, X_n$ are independent identically distributed random variables with $E[X_i]=m$ and $Var(X_i)=\sigma^2$ then for any $\epsilon > 0$,

$$P[|\frac{X_1+\cdots+X_n}{n} - m| \geq \epsilon] \to 0$$

We see this from Tchebychev's inequality.

$$P[|\frac{X_1+\cdots+X_n}{n} - m| \geq \epsilon] \leq \frac{Var(X_1+\cdots+X_n)}{n^2\epsilon^2} = \frac{Var(X)}{n\epsilon^2}$$

We need only $E[X]$ to exist. We can have $\mathcal{X}$ as an infinite set. $\sum_x p(x) = 1$ but $\sum_x F(x)p(x)$ does not converge. Geometric distribution. $p(x) = 2^{-x}$ for integers $x \geq 1$. If $X = F(x) = 2^x$ then $E[X] = \infty$.

Central Limit Theorem. Take coin tossing. n tosses. X is the number of heads. $\sigma(X) = \sqrt{\frac{n}{4}}$. What is the limit of

$$\Phi_n(z) = P\left[X \leq \frac{n}{2} + z\sqrt{\frac{n}{4}}\right]$$

as $n \to \infty$.

$$\lim_{n\to\infty} \Phi_n(z) = \Phi(z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}} du$$

$$\Phi_n(z) = \sum_{\frac{r-\frac{n}{2}}{\sqrt{\frac{n}{4}}} \leq z} \frac{n!}{r!(n-r)!} 2^{-n}$$

Stirling's approximation

$$n! \simeq \sqrt{2\pi} n^{n+\frac{1}{2}} e^{-n}$$

Riemann sum approximation. Actually if $\{X_i\}$ i.i.d.r.v with $E[X]=0$ and $Var(X_i)=\sigma^2$ then for $S_n = X_1 + \cdots + X_n$

$$P[S_n \leq \sqrt{n}\sigma z] \to \Phi(z)$$

One can use the moment generating function

$$E[e^{\theta X}] = M(\theta) = 1 + \theta E[X] + \frac{\theta^2}{2} E[X^2] + o(\theta^2)$$

If X, Y are independent

$$M_{X+Y}(\theta) = M_X(t) M_Y(t)$$

and

$$M_{cX}(t) = M_X(ct)$$

If $E[X] = 0$ and $E[X^2] = \sigma^2 = 1$, then

$$E[e^{\theta \frac{X_1 + \cdots + X_n}{\sqrt{n}}}] = [M(\frac{\theta}{\sqrt{n}})]^n \to e^{\frac{\theta^2}{2}} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{\theta x} e^{-\frac{x^2}{2}} dx$$

But it is better to use Fourier transforms or characteristic functions because $E[e^{\theta X}]$ may not be finite. The characteristic function is the analytic continuation of the moment generating function obtained by replacing θ with it.

$$\phi(t) = E[e^{itX}] = 1 + itE[X] - \frac{t^2}{2} E[X^2] + o(t^2)$$

If X, Y are independent

$$\phi_{X+Y}(t) = \phi_X(t) \phi_Y(t)$$

and

$$\phi_{cX}(t) = \phi_X(ct)$$

If $E[X] = 0$ and $E[X^2] = \sigma^2 = 1$

$$[\phi(\frac{t}{\sqrt{n}})]^n \to e^{-\frac{t^2}{2}} = \int e^{itx} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx$$

Sampling with and without replacement. CLT is valid if we sample with replacement. If we draw a sample of size n with replacement from a population of size N, the probability that we have no repetition is

$$\frac{N(N-1)\cdots(N-n+1)}{N^n} = \prod_{i=1}^{n-1} (1 - \frac{i}{N}) \simeq \exp[-\frac{n^2}{2N}]$$

If $\frac{n^2}{N}$ is small, then the central limit theorem is applicable even with replacement.

Stratified sampling.

If our population has two distinct groups, democrats and republicans and there are N_r and N_d of them with $N = N_r + N_d$ with M_r, M_d having favorable opinions with $M = M_r + M_d$, then to estimate $\frac{M}{N} = p$, it is better to estimate $p_r = \frac{M_r}{N_r}$ and $p_d = \frac{M_d}{N_d}$ separately and use $p = \frac{N_r}{N} p_r + \frac{N_d}{N} p_d$. It is best to split the sample size $n = n_r + n_d$ so that $\frac{n_r}{n} = \frac{N_r}{N}$.

Continuous distributions. If the space $\mathcal{X}$ is uncountable then probability distributions on it can not always specified by wiring down $\{p(x)\}$. For example one may want to think of a random variable uniformly distributed over the unit interval $[0,1]$. If $I = [a,b] \subset [0,1]$ then $P[I] = b - a$. More generally a probability distribution on the real line can be specified by a density $f(x)$ satisfying $f(x) \geq 0$ and $\int_{-\infty}^{\infty} f(x)dx = 1$. Then if X is a random variable having $f(x)$ as the density then

$$P[X \in A] = \int_A f(x)dx$$

and

$$E[H(X)] = \int_{-\infty}^{\infty} H(x)f(x)dx$$

If X is distributed with density $f(x)$ and $Y = g(X)$ is a smooth one to one function then the substitution $x = g^{-1}(y)$ converts

$$E[H(Y)] = \int H(g(x))f(x)dx = \int H(y)f(g^{-1}(y))\frac{dx}{dy}dy$$

so that Y is distributed with density $f(g^{-1}(y))\frac{1}{g'(g^{-1}(y))}$. If X is uniformly distributed on $[0,1]$ the distribution of $Y = X^2$ is given by the density

$$\frac{1}{2\sqrt{y}}$$

on $[0,1]$ (and 0 outside).

Classes of continuous distributions.

1. Normal family. $N(\mu, \sigma^2)$.

$$f_{\mu,\sigma}(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp[-\frac{(x-\mu)^2}{2\sigma^2}]$$

$$\int x f_{\mu,\sigma}(x) dx = \mu$$

and

$$\int (x-\mu)^2 f_{\mu,\sigma}(x) dx = \sigma^2$$

$$\int e^{itx} f_{\mu,\sigma}(x) dx = e^{it\mu - \frac{\sigma^2 t^2}{2}}$$

2. Gamma distributions.

$$f_{\alpha,p}(x) = \frac{\alpha^p}{\Gamma(p)} e^{-\alpha x} x^{p-1}$$

on $[0,\infty)$ and 0 otherwise.

3. Beta distributions.

$$f_{p,q}(x) = \frac{\Gamma(p+q)}{\Gamma(p)\Gamma(q)} x^{p-1}(1-x)^{q-1}$$

on $[0,1]$ and 0 otherwise.

Some facts.

If $\{X_i\}$ are independent with $X_i \simeq N(\mu_i, \sigma_i^2)$ then $Y = \sum_i X_i$ is distributed as $N(\sum_i \mu_i, \sum_i \sigma_i^2)$.

If $\{X_i\}$ are distributed as a Gamma with parameters $\{\alpha, p_i\}$, (a common value of α) then $Y = \sum_i X_i$ is distributed as Gamma with parameters $\alpha, \sum_i p_i$.

If X is $N(0,\sigma^2)$ then $Y = X^2$ is a Gamma with $\alpha = \frac{1}{2\sigma^2}, p = \frac{1}{2}$. In particular $\sum_{i=1}^n X_i^2$, the sum of squares of n independent normals with mean 0 variance 1 is Gamma with parameter $\alpha = \frac{1}{2}$ and $p = \frac{n}{2}$. This is called χ_n^2 or a chi-square with n degrees of freedom.

One can use characteristic functions. For the normal

$$E[e^{itX}] = \exp[it\mu - \frac{\sigma^2 t^2}{2}]$$

For the Gamma

$$E[e^{itX}] = \big(1 - \frac{it}{\alpha}\big)^{-p}$$

If X and Y are independent and Gamma distributed with parameters $(1, p)$ and $(1, q)$ respectively than $Z = \frac{X}{X+Y}$ has distribution that is Beta(p, q).

This is just the statement that

$$\begin{aligned} \frac{1}{\Gamma(p)\Gamma(q)} \int_0^\infty \int_0^\infty F(\frac{x}{x+y}) e^{-x-y} x^{p-1}(1-x)^{q-1} dx dy \\ = \frac{\Gamma(p+q)}{\Gamma(p)\Gamma(q)} \int_0^1 F(z) z^{p-1}(1-z)^{q-1} dz \end{aligned}$$

Change variables $z = \frac{x}{x+y}, u = x + y$ and integrate out u. We get $x = zu, y = (1-z)u$ and the Jacobian is

$$J = \begin{pmatrix} z & 1-z \\ u & -u \end{pmatrix}$$

$|J| = u$ and $dxdy = ududz$

$$\begin{aligned} \frac{1}{\Gamma(p)\Gamma(q)} \int_{-\infty}^\infty F(z) e^{-u} z^{p-1}(1-z)^{q-1} u^{p+q-1} du \\ = \frac{\Gamma(p+q)}{\Gamma(p)\Gamma(q)} z^{p-1}(1-z)^{q-1} \end{aligned}$$

Beta distribution of the first kind. On the other hand the distribution of $U = \frac{X}{Y} = \frac{Z}{1-Z}$ can be calculated by making the substitution $z = \frac{u}{1+u}$, $dz = \frac{du}{(1+u)^2}$.

$$\frac{\Gamma(p+q)}{\Gamma(p)\Gamma(q)} u^{p-1}(1+u)^{-(p+q)} du$$

is the density of Beta of the second kind. Some times one considers the ratio $F = \frac{\frac{X}{p}}{\frac{Y}{q}} = \frac{q}{p}U$. In other words $U = \frac{p}{q}F$. Note that $E[X] = p$ and $E[Y] = q$. The density of F is

$$\frac{\Gamma(p+q)}{\Gamma(p)\Gamma(q)}(\frac{p}{q})^p u^{p-1}(1+\frac{pF}{q})^{-p+q}$$

This is called the "F" distribution with (p,q) degrees of freedom or $F_{p,q}$ for short.

Another distribution that comes up is the t distribution, This the distribution of $\frac{X}{\sqrt{\frac{Y}{2p}}}$ where X is normal $N(0,1)$ and Y is a Gamma with parameters $(\frac{1}{2}, p)$ with $E[X] = 2p$. X and Y are assumed to be independent.

$$f(x)g(y) = \frac{1}{\sqrt{2\pi}}\frac{1}{\Gamma(p)}2^{-p}e^{-\frac{x^2}{2}}e^{-\frac{y}{2}}y^{p-1}dxdy$$

Make the substitutions $t = \frac{\sqrt{2p}x}{\sqrt{y}}$. $u = y$. Then $y = u$ and $x = \frac{t\sqrt{u}}{\sqrt{2p}}$.

$$J = \begin{pmatrix} \frac{\sqrt{u}}{\sqrt{2p}} & \frac{t}{2\sqrt{u}\sqrt{2p}} \\ 0 & 1 \end{pmatrix}$$

The Jacobian is $\frac{\sqrt{u}}{\sqrt{2p}}$. Need to calculate

$$c_p \int_0^\infty e^{-\frac{t^2 u}{4p}} e^{\frac{-u}{2}} u^{p-\frac{1}{2}} du = c'_p(1+\frac{t^2}{2p})^{-(p+\frac{1}{2})}$$

This is the density of the "t" distribution. The parameter $2p$ is the degrees of freedom. If p is large this is almost the normal density.

Multivariate distributions.

$$P[(X,Y) \in E] = \int_E h(x,y)dxdy$$

$$P[X \in A, Y \in B] = \int_A \int_B h(x,y)dxdy$$

If X and Y are independent have densities $f(x)$ and $g(y)$ then

$$h(x, y) = f(x)g(y)$$

When we say that if X and Y are independent and have densities f,g, then the density h of $X + Y$ is given by

$$\int\int F(x + y)f(x)g(y)dxdy = \int F(z)h(z)dz$$

$$h(z) = \int f(z - y)g(y)dy = \int g(z - x)f(x)dx$$

Sampling Distributions.

An independent sample of size n from a distribution with density $f(x)$ is just $X_1, \ldots, X_n$ having a distribution with a joint density $f(x_1) \cdots f(x_n)$ on R^n. A statistic is a function $t(X_1, X_2, \ldots, X_n)$ of the observations from $X_1, X_2, \ldots, X_n$. For example if we toss a coin n times and the probability of head is p, (a parameter that we do not know the value of) the observed proportion $t_n = \frac{\#(H)}{n}$ is a statistic. We saw before that $E[t_n] = p$ and Variance of t_n is $\frac{p(1-p)}{n}$. For large n, t_n will be close to p with very high probability. t_n is called a consistent estimate.

General parametric estimation:

We have a family $\{p(\theta, x)\}$ of probabilities or a family $p(\theta, x)$ of densities. We have n independent observations $X_1, \ldots, X_n$. Estimate θ. $\hat{\theta} = t(X_1, \ldots, X_n)$.

Unbiased is good. $E_\theta[t_n(X_1, X_2, \ldots, X_n)] = \theta]$ for all θ.

Must control variance as well.

$$E_\theta[[t_n(X_1, \ldots, X_n) - \theta)]^2]$$

is small.

Statistics from normal random variables.

Sum of n independent normal variables is again normal with mean and variance equal to the sum of the means and variances.

Sum of squares of n normal random variables with mean 0 variance 1 is called χ_n^2 a chi-square with n degrees of freedom.

The sample mean is always an unbiased estimator of the population mean because

$$E[\frac{X_1 + \cdots + X_n}{n}] = \frac{1}{n}\sum_{i=1}^{n} E[X_i] = E[X]$$

Sample variance:

$$s^2 = \frac{1}{n}\sum(x_i - \bar{x})^2 = \frac{1}{n}\sum x_i^2 - \bar{x}^2$$

$$E[s^2] = \frac{n-1}{n}\sigma^2$$

Orthogonal change of variables. $y = Tx$. $y_1 = \sqrt{n}\bar{x}$. Choose the remaining coordinates so that $y_2, \ldots, y_n$ to form an orthonormal basis so that T is an orthogonal matrix.

$$ns^2 = \sum x_i^2 - n\bar{x}^2 = \sum_{i=1}^{n} y_i^2 - y_1^2 = \sum_{i=2}^{n} y_i^2$$

$$\bar{x} = \frac{y_1}{\sqrt{n}}$$

$$ns^2 = \sum x_i^2 - n\bar{x}^2 = \sum_{i=1}^{n} y_i^2 - y_1^2 = \sum_{i=2}^{n} y_i^2$$

$$\bar{x} = \frac{y_1}{\sqrt{n}}$$

$\frac{ns^2}{n-1}$ is a good estimator of σ^2

$$t = \frac{\sqrt{n}\bar{x}}{\sqrt{\frac{ns^2}{n-1}}} = \frac{\bar{x}}{s}\sqrt{n-1}$$

is a t with $n-1$ degrees of freedom.